Steve Depp
413 – 55

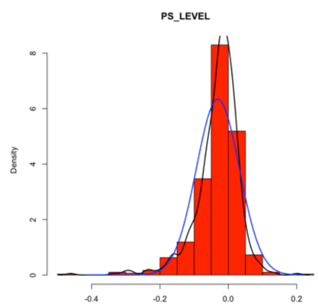
Assignment 5
22 July 2019

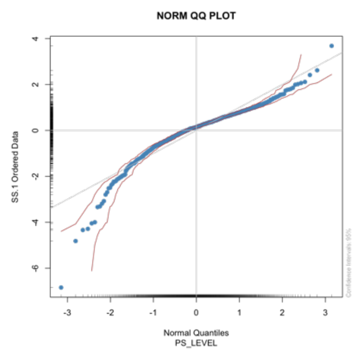
1.1 Perform EDA.

The following presents sample size, mean, standard deviation, median, 20% trimmed mean, min, max, range, skewness, excess kurtosis for levels of liquidity, PS_LEVEL. EDA first confirms there are no gaps in the 605 monthly observations and converts dates from YYYYMMDD format to monthly percent of year format.

| | n | mean | sd | median | trimmed | min | max | range | skew | kurtosis |
|---|---|---|---|---|---|---|---|---|---|---|
| **X1** | 605 | -0.03063832 | 0.06290012 | -0.02152052 | -0.02493282 | -0.4610066 | 0.2010149 | 0.6620215 | -1.529748 | 6.265555 |

The **above** univariates describe a distribution departing from normal via negative skew, and positive kurtosis. The -0.13 midpoint of the range is < mean < trimmed mean < median implying that although the distribution stretches far to the left, and the middle 60% are skewed left, many observations are to the right.

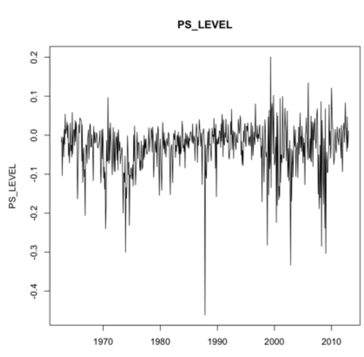


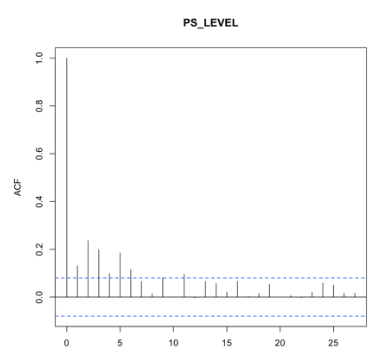



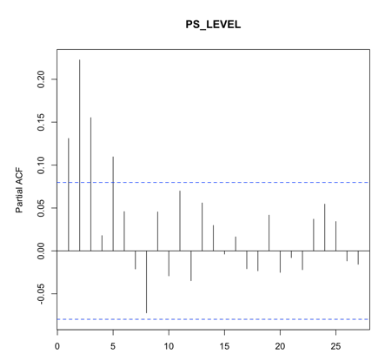

```
[1] "H1: mean=zero"

        One Sample t-test

data:  xt_1
t = -11.981, df = 604, p-value < 2.2e-16
alternative hypothesis: true mean is not equal to 0
95 percent confidence interval:
 -0.03566051 -0.02561614
sample estimates:
  mean of x
-0.03063832
```

The p-value = 2.2e-16 from a one sample t-test allows us to **reject** the hypothesis that the distribution has a **zero mean** at the highest level of confidence, i.e. > 99.9% as shown **above** on the right. Similarly, a test of null hypotheses that the distribution is symmetric with respect to the mean yields a test statistic = -15.36 and p-value ~zero, allowing us to reject the null hypothesis of zero skewness with greater than 99.9% level of confidence. A test for kurtosis yields a test statistic 31.46 and p-value ~zero allowing us to reject the null hypothesis of zero kurtosis with greater than a 99.9% level of confidence.

The **above** histogram and density plot overlaid normal curve represent the described left skewed leptokurtotic distribution. The negative skew / fat left tail can be seen in the **above** QQ plot as (a) data minimums extend to ~7 SDs versus maxima to ~4 SDs. Outliers number 5 between 4-6 SDs on the left plus the one at 7SDs versus a single outlier at ~3.7 SDs on the right.




The time series plot **above** appears to have a constant mean close to zero. 12 or more spikes appear to extend more than 0.20 below the -0.03 mean, including the 5 mentioned above with values -0.28 to -0.33 or roughly 5 times the 0.063 sample distribution standard deviation and 1 at the -0.46 = ~7 SDs from the mean. These points illustrate the univariate negative skew and kurtosis measurements. Variance is not constant with areas of greater turbulence noted twice on the right-hand side of the plot and once on the left. There's no discernable seasonality or trend or cyclicality observable from this dense plot.

```
        Box-Ljung test

data:  xt_1
X-squared = 103.34, df = 6, p-value < 2.2e-16
```

As measured by 95% confidence intervals in blue, the ACF and PACF plots **above** contain autocorrelations most prominent at lag 2, 3 and 5 but significant out to lag 6 which might support an MA model and significant but varying levels of partial autocorrelations at 1, 2, 3, and 5 which might support an AR model at those lags. Consistent with the ACF/PACF plots, the Box-Ljung portmanteau test of autocorrelation shown **above** has a p-value ~zero suggesting correlation inside lag 6.

```
Title:
 Augmented Dickey-Fuller Test

Test Results:
  PARAMETER:
    Lag Order: 6
  STATISTIC:
    Dickey-Fuller: -6.6464
  P VALUE:
    0.01
```

```
#######################
# KPSS Unit Root Test #
#######################

Test is of type: mu with 6 lags.

Value of test-statistic is: 0.3378

Critical value for a significance level of:
                10pct  5pct 2.5pct  1pct
critical values 0.347 0.463  0.574 0.739
```

There are several tests for unit roots to apply to this data. **Firstly**, the Augmented Dickey Fuller Test's null hypothesis is that there are unit roots and the data is not stationary. The alternate hypotheses are that the series is stationary or trend stationary with a constant but no trend (type = "c") that appears to fit our raw data. Testing the null hypothesis with a lag = 6 **above** yields a test statistic = -6.6464 which allows us to reject the null hypothesis that our data is non-stationary and has unit roots with greater than 99.9% confidence. **Secondly**, KPSS Unit Root Test automatically employs lag = 6 for our data set and tests the null hypothesis that our data are stationary yielding a test statistic = 0.3378 which is less than 0.347 necessary to reject the null hypothesis at the 90% confidence level. Thus we cannot reject the null hypothesis. **Finally**, KPSS tests repeatedly executed via the 'ndiffs' function to determine the appropriate number of first differences, which in this case is zero, indicating no differencing is needed.

```
Series: xt_1
ARIMA(3,0,3) with non-zero mean

Coefficients:
          ar1     ar2     ar3     ma1     ma2      ma3     mean
      -0.0523  0.1586  0.5159  0.1116  0.0443  -0.3596  -0.0304
s.e.   0.3981  0.2792  0.1482  0.4065  0.2756   0.1395   0.0051

sigma^2 estimated as 0.003602:  log likelihood=846.88
AIC=-1677.75   AICc=-1677.51   BIC=-1642.51
```

```
Series: xt_1
ARIMA(5,0,5) with non-zero mean

Coefficients:
         ar1     ar2      ar3     ar4     ar5      ma1     ma2     ma3      ma4
      0.2867  0.0112  -0.2521  0.3638  0.0011  -0.2243  0.1620  0.3579  -0.3469
s.e.  0.2757  0.2571   0.1931  0.2258  0.2241   0.2730  0.2541  0.1902   0.1722
         ma5     mean
      0.1998  -0.0306
s.e.  0.1966   0.0047

sigma^2 estimated as 0.00358:  log likelihood=850.74
AIC=-1677.48   AICc=-1676.95   BIC=-1624.62
```

```
Series: xt_1
ARIMA(3,0,2) with non-zero mean

Coefficients:
          ar1     ar2     ar3     ma1      ma2     mean
      -0.3034  0.6586  0.1993  0.3643  -0.4636  -0.0305
s.e.   0.1158  0.0860  0.0493  0.1161   0.1085   0.0049

sigma^2 estimated as 0.003588:  log likelihood=847.57
AIC=-1681.13   AICc=-1680.95   BIC=-1650.3
```

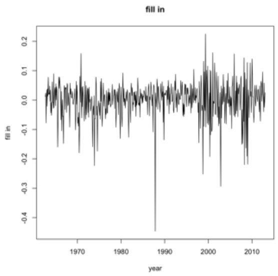
```
Call:
arima(x = da_1[, 2], order = c(5, 0, 0))

Coefficients:
         ar1     ar2     ar3     ar4     ar5  intercept
      0.0624  0.1858  0.1313  0.0105  0.1096    -0.0306
s.e.  0.0404  0.0404  0.0408  0.0405  0.0404     0.0048

sigma^2 estimated as 0.003555:  log likelihood = 847.3,  aic = -1680.6
```

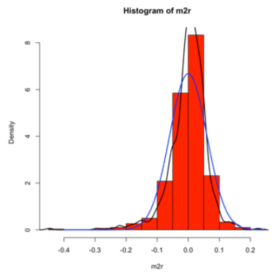
1.2 Experimenting with ARMA models (3,0,3) **top left above** and (5,0,5) **top right above** yield AIC levels ~-1677.5 though coefficients are mostly not significant. Employing auto.arima and setting stepwise and approximation to FALSE yields an ARMA model (3,0,2) **bottom left above** with AIC=-1681.13. Equation of model to be fitted is:

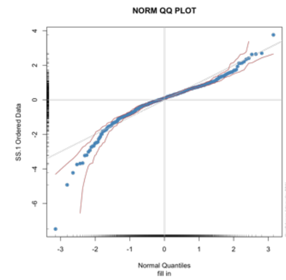
$$(1-\varphi_1 B^1-\varphi_2 B^2-\varphi_3 B^3)\, y_t = c+(1+\theta_1 B^1+\theta_2 B^2)\varepsilon_t \text{ or } \phi(\mathrm{B})\, y_t = c+\theta(\mathrm{B})\varepsilon_t$$

The model of order (5,0,0) shown **bottom right above** would have the equation **below**:

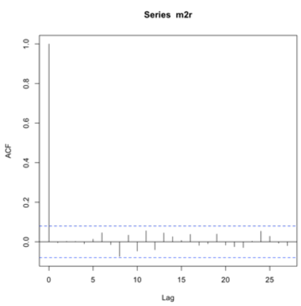
$$(1-\varphi_1 B^1-\varphi_2 B^2-\varphi_3 B^3-\varphi_4 B^4-\varphi_5 B^5)\, y_t = c + \varepsilon_t$$

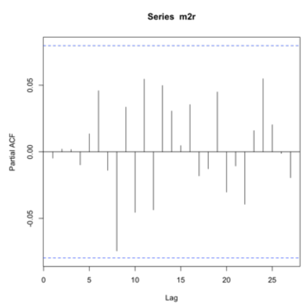







The residuals of the requested (5,0,0) model which are similar to those of the auto-arima (3,0,2) model, with a zero mean whose null hypothesis cannot be rejected and levels of negative skew and leptokurtosis significant at 99.9% level. The time series, histogram, density and QQ plots **above** are virtually identical to that of the dependent variable discussed above, skewed to 7 SDs on the left with 5 outliers to 5 SDs plus 1 to 7 SDs and 1 outlier on the right to nearly 4 SDs.






```
        Box-Ljung test

data:  m2r
X-squared = 1.4747, df = 6, p-value = 0.9612
```

ACF/PACF show stationary times series. Box-Ljung confirms independence to 6 lags.

```
Call:
arima(x = da_1[, 2], order = c(5, 0, 0), xreg = i303)

Coefficients:
         ar1     ar2     ar3     ar4     ar5  intercept     i303
      0.0657  0.1884  0.1309  0.0162  0.1261    -0.0299  -0.4255
s.e.  0.0403  0.0404  0.0408  0.0404  0.0403     0.0049   0.0550

sigma^2 estimated as 0.003236:  log likelihood = 875.8,  aic = -1735.59
```

1.3 The largest outlier at the 303rd position with a model residual value 7.48 SDs below the mean, as observed in the plots above. Fitting the residuals from the model with order (5,0,0) to the i303 binary regressor containing all zeros except this at the 303rd position where the regressor value is set = 1 obtains the above model and this model equation:

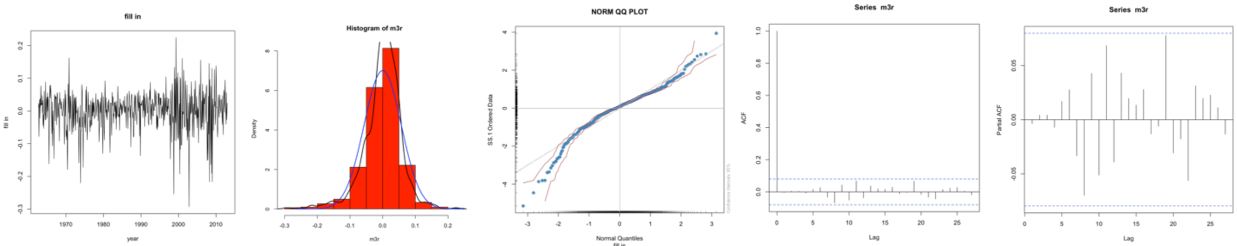
$$(1-0.0657*B^1-0.1884*B^2-0.1309*B^3-0.0162*B^4-0.1261*B^5)\, y_t = -0.0299 -0.4255*i303$$

| | n | mean | sd | median | trimmed | min | max | range | skew | kurtosis |
|---|---|---|---|---|---|---|---|---|---|---|
| X1 | 605 | 2.956233e-05 | 0.05967631 | 0.005517134 | 0.004047833 | -0.4462932 | 0.2247652 | 0.6710584 | -1.427216 | 7.403997 |

| | n | mean | sd | median | trimmed | min | max | range | skew | kurtosis |
|---|---|---|---|---|---|---|---|---|---|---|
| X1 | 605 | 2.308778e-05 | 0.05692922 | 0.004306856 | 0.003358947 | -0.2930509 | 0.2241593 | 0.5172101 | -0.8572538 | 3.295622 |

Before and after comparisons of residual distributions from the 2 models **above** show that negative skew is cut by 40% from -1.43 to -0.86 and leptokurtosis by 55% from 7.4

to 3.3, but the mean remains zero and skew and leptokurtosis significant at the 99.9% confidence level.

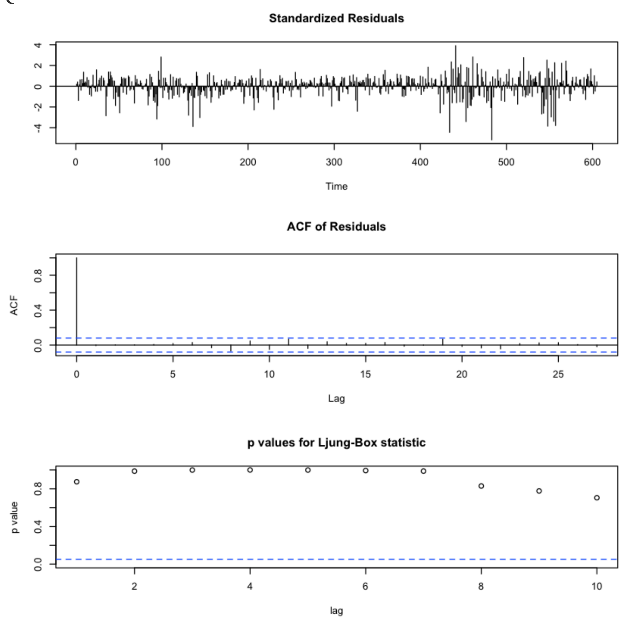



Time series plots, histograms, densities, QQ, ACF, PACF are all largely unchanged excepting the removal of the outlier as shown **above**.

```
Call:
arima(x = da_1[, 2], order = c(5, 0, 0), xreg = i303, fixed = c1)

Coefficients:
         ar1     ar2     ar3  ar4     ar5  intercept     i303
      0.0683  0.1917  0.1318    0  0.1272    -0.0299  -0.4248
s.e.  0.0398  0.0395  0.0407    0  0.0402     0.0048   0.0549

sigma^2 estimated as 0.003236:  log likelihood = 875.71,  aic = -1737.43
```

1.4 Refining the model by zeroing the 4th and least significant parameter (t-ratio = 0.26) obtains the model **above** with AIC improved from -1735.59 to 1737.43. Model equation:

$$(1-0.0683*B^1-0.1917*B^2-0.1318*B^3-0*B^4-0.1272*B^5)\, y_t = -0.0299 -0.4248*i303$$




Residuals continue stationary and independent as shown **above**. Other residual distribution plots are largely unchanged, but not shown here.

```
Title:
 Augmented Dickey-Fuller Test

Test Results:
  PARAMETER:
    Lag Order: 5
  STATISTIC:
    Dickey-Fuller: -9.464
  P VALUE:
    0.01
```

```
#######################
# KPSS Unit Root Test #
#######################

Test is of type: mu with 6 lags.

Value of test-statistic is: 0.2118

Critical value for a significance level of:
                10pct  5pct 2.5pct  1pct
critical values 0.347 0.463  0.574 0.739
```

```
        Box-Ljung test

data:  m3.2r
X-squared = 0.76028, df = 6, p-value = 0.9931
```

As shown **above**, ADF & KPSS confirm no unit roots to 5 & 6 lags.
Box-Ljung confirms independence to lag 6.

2.1 Perform EDA.

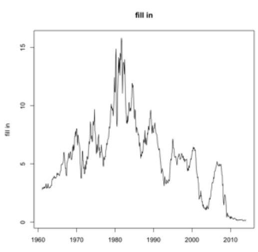
The following presents sample size, mean, standard deviation, median, 20% trimmed mean, min, max, range, skewness, excess kurtosis for levels of liquidity, PS_LEVEL. EDA first confirms there are no gaps in the 636 monthly observations and converts dates from YYYYMMDD format to monthly percent of year format.

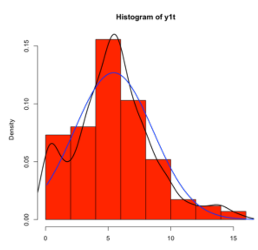
| | n | mean | sd | median | trimmed | min | max | range | skew | kurtosis |
|---|---|---|---|---|---|---|---|---|---|---|
| **y1t** | 636 | 5.434388 | 3.144130 | 5.403 | 5.296794 | 0.106 | 15.812 | 15.706 | 0.5033152 | 0.4764783 |
| **y3t** | 636 | 5.825406 | 3.004899 | 5.708 | 5.693743 | 0.290 | 15.557 | 15.267 | 0.4611179 | 0.3212491 |

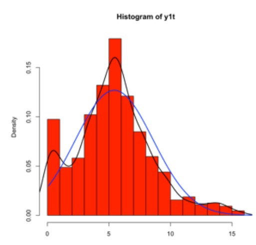
```
        One Sample t-test

data:  y1t
t = 43.589, df = 635, p-value < 2.2e-16
alternative hypothesis: true mean is not equal to 0
95 percent confidence interval:
 5.189567 5.679209
sample estimates:
mean of x
 5.434388
```

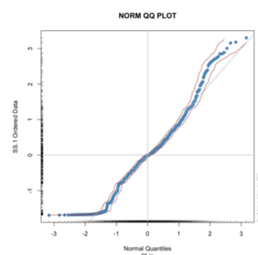
```
        One Sample t-test

data:  y3t
t = 48.891, df = 635, p-value < 2.2e-16
alternative hypothesis: true mean is not equal to 0
95 percent confidence interval:
 5.591426 6.059385
sample estimates:
mean of x
 5.825406
```

The **above** univariates describe a distribution departing from normal via moderately positive skew, and kurtosis. The trimmed mean < median < mean < midpoint of range for both series implying that the core of the distribution is skewed left while the overall distribution is skewed right with tails that extend far to the right. T-tests of one sample handily reject the null hypothesis that the mean of both distributions is zero. Similarly, at a confidence level of 99.9% one can reject the hypothesis that both distributions' skew is zero. A null hypothesis test of kurtosis = 0 can be rejected at the 98.5% level for 1y yields and at 90% for 3y yields (test statistic = 2.45 and 1.65 respectively).

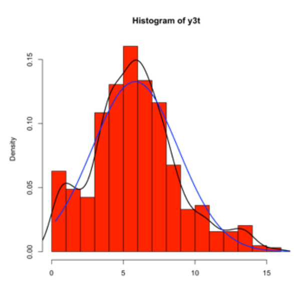


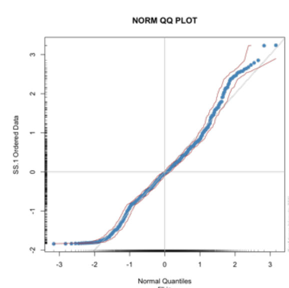



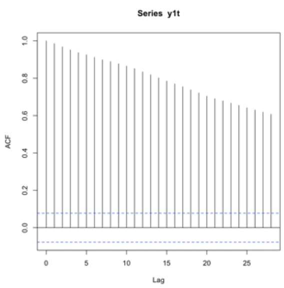



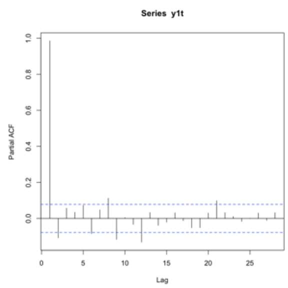

From time series plot of 1y yields **above**, one observes no constant mean, up cycles of varying lengths, 2 down cycles on the left of similar length, 3 down cycles on the right of similar length, and different variances in different ranges of the plot. The 2nd histogram is more granular in order to see the frequency of observations in the low yields. It's possible these are the values dragging the mean lower. Otherwise, the distribution appears positively skewed or very positively skewed. Real interest rates during this period of low yields might have been negative, and in a sense therefore, this distribution is truncated where the high bar shows on the far left. The QQ plot bears this out with many observations that normally would be 1.5-3.0 SDs from the mean truncated at ~1.8 SDs from the mean on the LHS. Also visible on the QQ are fat tails on the RHS.

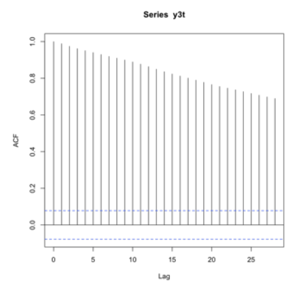



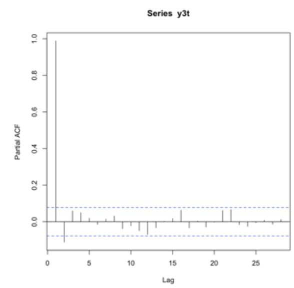



Plots of time series, histograms, densities, and QQ for 3 year yields **above** are very similar to those observed for 1y yields. The lengths of the down cycles on the LHS and RHS are similar again to each other though here the depths of declines appear more shallow. The bar on the LHS of this histogram is not as prominent as for 1y yields which might explain the lower leptokurtosis of 3 year yields. The QQ plot shows similar truncation as for 1y yields but at a level closer to 2 SD than was observed for 1y yields. There are also fewer outliers, 2 versus 3 or 4 in this QQ plot of 3 year than observed above in 1 year yields.










Both PACF plots show strong partial autocorrelation at lag 1 which explains the steady decline in autocorrelation in ACF plots. Also evident are significant PACF at lags 2, 6, 8, 9, 12 for 1y yields and lag 2 for 3-year yields. Any 6 of these could be due to the 95% confidence level employed by the confidence band shown. So further investigation via models will be needed.

```
Title:
 Augmented Dickey-Fuller Test

Test Results:
  PARAMETER:
    Lag Order: 5
  STATISTIC:
    Dickey-Fuller: -2.3741
  P VALUE:
    0.4199
```

```
Title:
 Augmented Dickey-Fuller Test

Test Results:
  PARAMETER:
    Lag Order: 5
  STATISTIC:
    Dickey-Fuller: -1.2646
  P VALUE:
    0.588
```

```
#######################
# KPSS Unit Root Test #
#######################

Test is of type: mu with 6 lags.

Value of test-statistic is: 3.0372

Critical value for a significance level of:
                10pct  5pct 2.5pct  1pct
critical values 0.347 0.463  0.574 0.739
```

```
#######################
# KPSS Unit Root Test #
#######################

Test is of type: mu with 6 lags.

Value of test-statistic is: 3.0369

Critical value for a significance level of:
                10pct  5pct 2.5pct  1pct
critical values 0.347 0.463  0.574 0.739
```

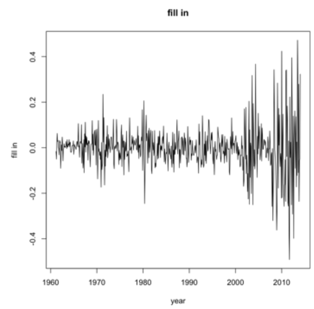
Tests for unit roots applied to this data confirm lag 1 autocorrelation observed in the PACF. **Firstly**, the Augmented Dickey Fuller Test's null hypothesis is that there are unit roots and the data is not stationary. The alternate hypotheses are that the series is stationary or trend stationary with a constant but no trend (type = "c") that appears to fit our raw data. Testing the null hypothesis with a lag = 5 **above** yields test statistics = -2.37 for 3y and -1.26 for 1y yields with pvalues 0.42 and 0.59 which prevent us from rejecting the null hypothesis that our data is non-stationary and has unit. **Secondly**, KPSS Unit Root Test automatically employs lag = 6 for our data set and tests the null hypothesis that our data are stationary yielding a test statistic = 3.04 for both series which is greater than 0.74 critical value for rejecting the null hypothesis at the 99% confidence level. **Finally**, KPSS tests repeatedly executed via the 'ndiffs' function to determine the appropriate number of first differences, which in this case is 1, a first differencing is needed.

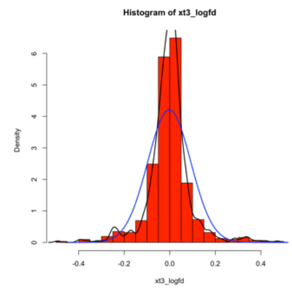
To rectify the skewed distribution and unit roots, the 3y time series is log transformed and 1$^{st}$ differenced.

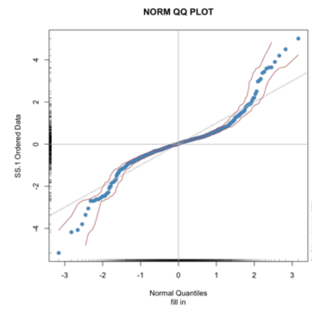
| | n | mean | sd | median | trimmed | min | max | range | skew | kurtosis |
|---|---|---|---|---|---|---|---|---|---|---|
| X1 | 635 | -0.002254463 | 0.09474429 | -0.0007430341 | -0.002085373 | -0.4918227 | 0.4723229 | 0.9641456 | 0.1161441 | 5.66895 |

```
        One Sample t-test

data:  xt3_logfd
t = -0.59962, df = 634, p-value = 0.549
alternative hypothesis: true mean is not equal to 0
95 percent confidence interval:
 -0.009637654  0.005128728
sample estimates:
   mean of x
-0.002254463
```

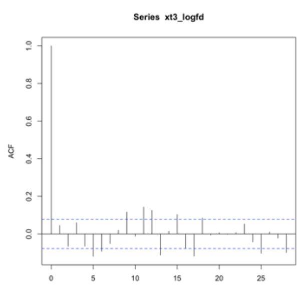
The resulting distribution univariates shown **above** a lot more leptokurtosis and a great reduction in positive skew. The mean is affirmed zero by p-value = 0.55 meaning we cannot reject the null hypothesis that mean = 0. While, we can reject the null hypothesis of zero excess kurtosis at a confidence level of greater than 99.9%, the test statistic = 1.19 associated skew prevents us from rejecting the null hypothesis of zero skew.

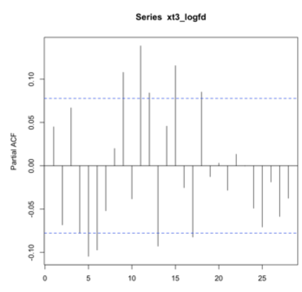



Plots **above** bear out the conclusions of zero constant mean, a much more symmetric distribution, significant leptokurtosis mostly due to thin shoulders and a few outliers on both sides. Otherwiswe, the QQ plot represents most of the distribition inside of 3 SDs on both sides. On the LHS, there are 6 outliers extending to 4.8 SDs and on the RHS there are 9 extending to about 4.8 SDs as well. Notable is the expansion of variance in the preiod from about 2003-2006 and then even more prominently from 2008 onwards where ranges are double in the first case and easily quadruple in the second case when compared with the years prior to 2003. These are the years associated with the outliers on both ends of the distribution.




The ACF and PACF plots show greatly changed autocorrelation, signficant in ACF lags 5, 9, 11, 12, 13, 15, 16, 17, 18 and for PACF lags 5, 6, 9, 11, 13, 15, 17, 18.

```
Title:
 Augmented Dickey-Fuller Test

Test Results:
  PARAMETER:
    Lag Order: 5
  STATISTIC:
    Dickey-Fuller: -12.6005
  P VALUE:
    0.01
```

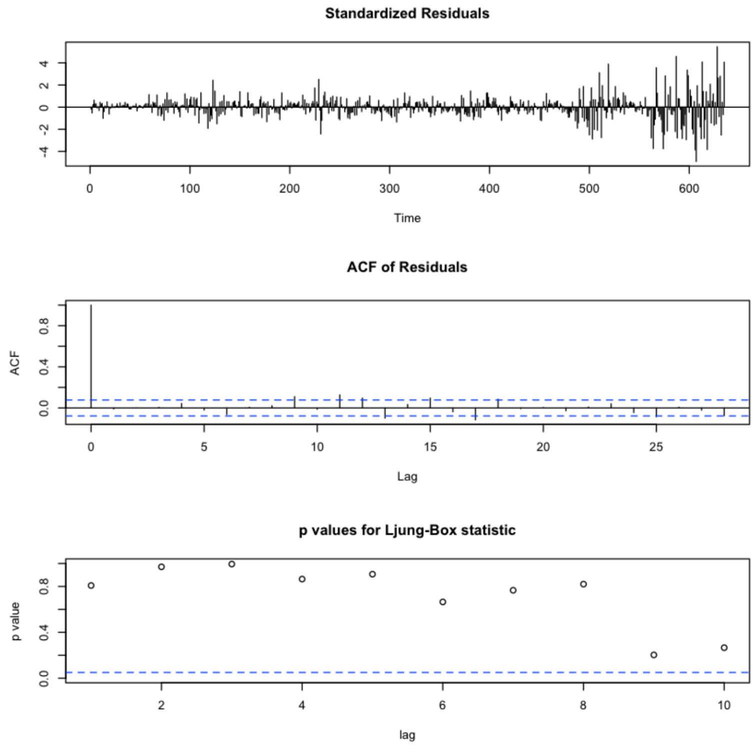
```
#######################
# KPSS Unit Root Test #
#######################

Test is of type: mu with 6 lags.

Value of test-statistic is: 0.3035

Critical value for a significance level of:
                10pct  5pct 2.5pct  1pct
critical values 0.347 0.463  0.574 0.739
```

Switching to no constant / no trend ADF testing enables us to reject the hypothesis of non-stationary series with unit roots at the 99% level. KPSS Unit Root test with 6 lags bears this result as well with its 0.3035 test statistic falling below the 90% confidence level. Thus it is concluded that no unit roots exist for the 1st difference of log time series of 3 year yields. It is suspected but not confirmed that the same is true for 1y yields.

```
Series: xt3_logfd
ARIMA(4,0,1) with zero mean

Coefficients:
         ar1      ar2     ar3      ar4      ma1
      0.6369  -0.1116  0.1169  -0.1515  -0.5858
s.e.  0.1124   0.0472  0.0474   0.0396   0.1086

sigma^2 estimated as 0.008767:  log likelihood=605.33
AIC=-1198.67   AICc=-1198.53   BIC=-1171.95
```

$(1-\phi_1 B - \phi_2 B^2 - \phi_3 B^3 - \phi_4 B^4)(1-B)\ln(y_t) = (1+\theta_1 B)\varepsilon_t$
$(1-0.6369B - (-0.1116)B^2 - 0.1169B^3 - (-0.1515)B^4)(1-B)\ln(y_t) = (1+(-0.5858)B)$

2.2 The auto.arima() function with stepwise and approximation set to FALSE produces the model **above** with 4 AR and 1 MA.

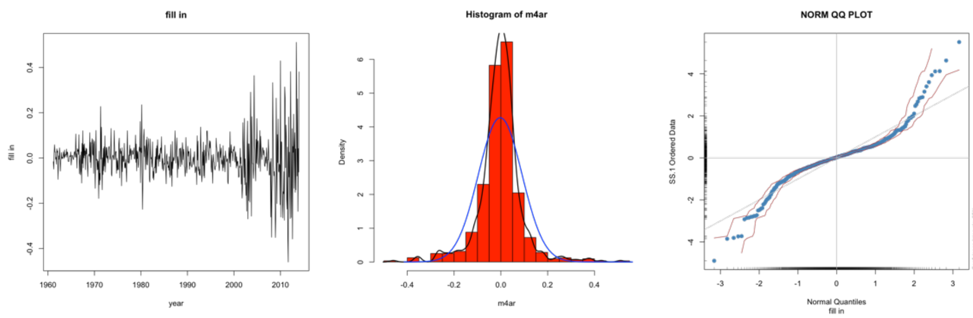



Cursory examination of the model residuals above shows similar variance expansion on the RHS. Otherwise the ACF appear to represent a stationary time series of residuals and p-values for Ljung Box an independent series of residuals.

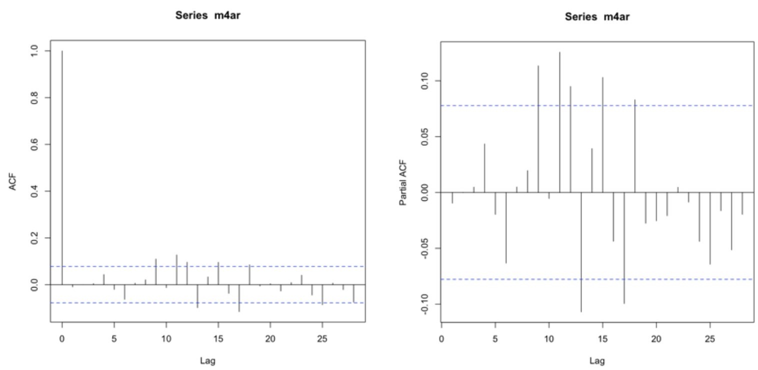
```
        One Sample t-test

data:  m4ar
t = -0.76931, df = 634, p-value = 0.442
alternative hypothesis: true mean is not equal to 0
95 percent confidence interval:
 -0.010118475  0.004422055
sample estimates:
  mean of x
-0.00284821
```

| | n | mean | sd | median | trimmed | min | max | range | skew | kurtosis |
|---|---|---|---|---|---|---|---|---|---|---|
| X1 | 635 | -0.00284821 | 0.09329517 | 0.0001443481 | -0.001872323 | -0.4595939 | 0.5113095 | 0.9709034 | 0.1541066 | 5.684125 |

The residuals' univariate statistics represent a distribution similar to the original 1st differenced log series: leptokurtotic with mild skew and a zero mean. Hypothesis testing of zero mean does not enable rejection of a zero mean (pvalue = 0.442). Likewise, we affirm a symmetric distribution via inability to reject a test of the null hypothesis of zero skew (pvalue = .113). With a test statistic of 29.23, we can reject the null hypothesis of zero excess kurtosis at above the 99.9% confidence level.




Similar to the distribution of the log 1st differenced series, the residuals of this ARIMA model exhibit variance expanding sharply toward the RHS of the time series, a peaked leptokurotic histogram, and residual outliers that extend well into the 4th SD on both sides of the QQ plot.




ACFs are independent, but show significant correlation in lag 9, 11, 12, 13, 15, 17. PACF of model residuals exhibit significant autocorrelation at lag 9, 11, 12, 13, 15, 17, 18.

```
        Box-Ljung test

data:  m4ar
X-squared = 4.0867, df = 6, p-value = 0.6649
```

```
        Box-Ljung test

data:  m4ar
X-squared = 42.762, df = 16, p-value = 0.0003034
```

Portmaneau test of autocorrelation inside lag 6 does not allow us to reject the null hypothesis of a stationary series, but from lag 16 onwards, the test can be rejected at a confidence level of 99.9%. Unit root tests (ADF and KMSS) were run and reject unit roots for these residuals.

OUTLIERS: This original series and model residual series represent outliers which appear to be concentrated in a period of time mentioned earlier 2003-6 and 2008 onwards. Inasmuch as sampling error is unlikely the cause, and these observations are part of an expansion in variance in the time series, it wouldn't be wise to delete the observations or even impute them. It is possible an indicator associated with the outliers predicts other market prices but more likely there is a different model present during these years. Likely a good test would be to run models for these years and compare with models from previous years. If the models are adequate but different, there may be value in separating the modeling process completely.

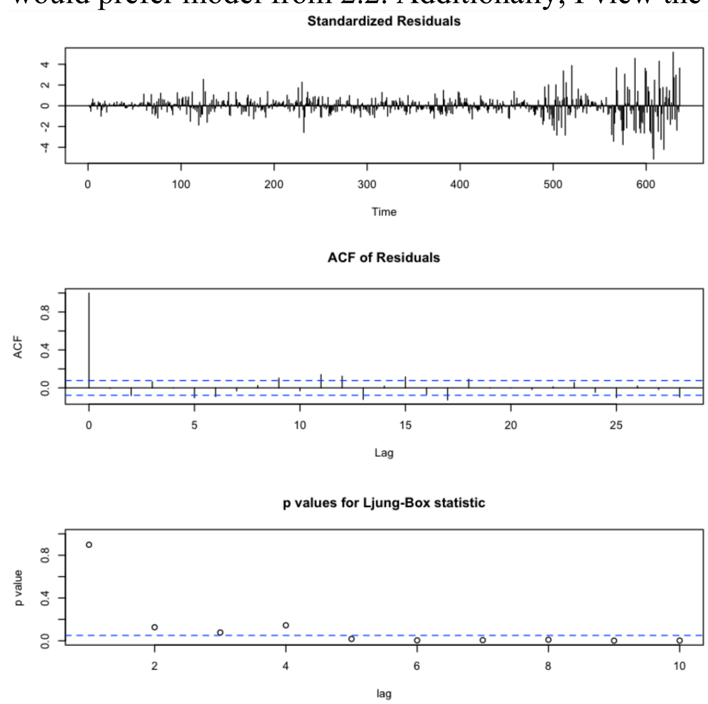
2.3 Taking the log of 3y yields and employing the model called for in the instructions with order = c(0,1,1) and seasonal=list(order=c(0,0,1), period=4) will 1st difference the log 3 year yields and so the modeled data should be the same. The model will look for 1 period moving average and 1 seasonal moving average of 4 periods per year or quarterly.

```
Call:
arima(x = xt3_log, order = c(0, 1, 1), seasonal = list(order = c(0, 0, 1), period = 4))

Coefficients:
         ma1     sma1
      0.0533  -0.0617
s.e.  0.0441   0.0385

sigma^2 estimated as 0.00891:  log likelihood = 597.76,  aic = -1189.52
```

$(1-B)\ln(y_t)=(1+\theta_1 B)\ (1+\Theta_1 B_3)\varepsilon_t.$

$(1-B)\ln(y_t)=(1+0.0533*B)\ (1+(-0.0617)*B_3)$

Nominally speaking this model's AIC is -1189.52 versus 2.2 model AIC -1198.67. So, I would prefer model from 2.2. Additionally, I view the cursory plots below:




While the residuals appear to be zero meaned and stationary from the time series and ACF plots **above**, the Box-Ljung pvalues are all very close to zero, indicating that we have left a lot of autocorrelation between periods unexplained in this model.

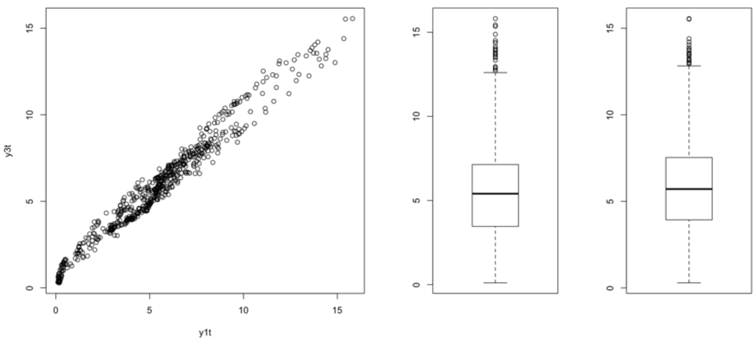
2.4 Employing backtest.R on these models yields RMSE 0.2274 for the model in problem 2.2 and RMSE of 0.2309 for the model here in problem 2.3. With a lower RMSE, I would prefer the model from problem 2.2.

3.1 Measure time series correlations and significance below:

```
0.984408679254895

        Kendall's rank correlation tau

data:  y1t and y3t
z = 33.741, p-value < 2.2e-16
alternative hypothesis: true tau is not equal to 0
sample estimates:
      tau
0.8945081
```

```
0.751223631651215

        Kendall's rank correlation tau

data:  xt1_logfd and xt3_logfd
z = 24.206, p-value < 2.2e-16
alternative hypothesis: true tau is not equal to 0
sample estimates:
      tau
0.6421702
```

For nominal yields 1y vs 3y on the LHS **above** and for the 1st difference of their logs on the RHS above demonstrates stronger bivariate association between nominal yields than between their 1st differenced logs. These results produce a test value that allow us to reject the hypothesis that correlations are zero in both cases at greater than a 99.9% confidence level.




The plots **above** represent bivariate scatter of nominal yields, y1t versus y3t and box plots demonstrating a close association. The association is closer at lower yields than at higher yields in the scatter plot where the association is dense on the LHS and sparse on the RHS.

$y3_t = \beta_0 + \beta_1 * y1_t + e_t$

3.2 A linear regression of the 3 year yields, y3t, as the dependent variable against 1 year yields, y1t as independent variable produces this model equation **above** to be fit:

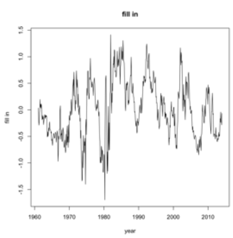
```
Call:
lm(formula = y3t ~ y1t)

Residuals:
     Min       1Q   Median       3Q      Max 
-1.69599 -0.42038 -0.03045  0.37993  1.41445 

Coefficients:
            Estimate Std. Error t value Pr(>|t|)    
(Intercept) 0.712645   0.041909    17.0   <2e-16 ***
y1t         0.940816   0.006676   140.9   <2e-16 ***
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

Residual standard error: 0.529 on 634 degrees of freedom
Multiple R-squared:  0.9691,	Adjusted R-squared:  0.969 
F-statistic: 1.986e+04 on 1 and 634 DF,  p-value: < 2.2e-16
```

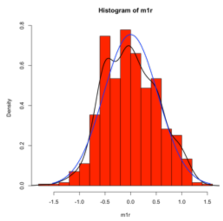
According to the model output above, the model is significant with p-value for the model and its explanatory parameters of <2e-16 or zero, enabling us to reject the hypothesis that the parameter is zero. R-squared is 0.969 which suggests a good fit.

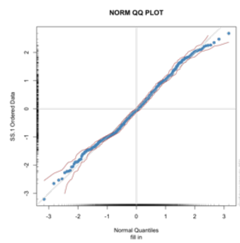
| | n | mean | sd | median | trimmed | min | max | range | skew | kurtosis |
|---|---|---|---|---|---|---|---|---|---|---|
| X1 | 636 | -2.218451e-17 | 0.528551 | -0.03044947 | -0.01144073 | -1.695991 | 1.414454 | 3.110445 | 0.1167925 | -0.3985518 |

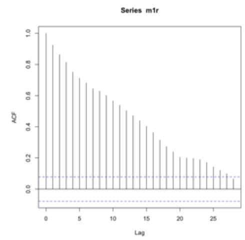
```
	One Sample t-test

data:  mlr
t = -1.0585e-15, df = 635, p-value = 1
alternative hypothesis: true mean is not equal to 0
95 percent confidence interval:
 -0.04115617  0.04115617
sample estimates:
    mean of x 
-2.218451e-17 
```

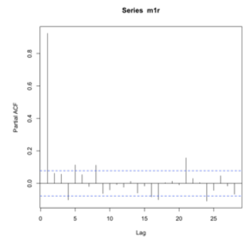
The mean of residuals = 0 with a p-value = 1 suggesting we cannot reject the null hypothesis that the mean = zero. Likewise, the skewness of the residuals tests with a p-value = 0.23 suggesting we cannot reject the null hypothesis of zero skew. Kurtosis is likely non-zero with a confidence leel of 0.95% that we can reject the null hypotheis of zero kurtosis.












The time series, histogram, density plot, QQ, ACF and PACF suggest that the residuals are normally distributed and stationary but that they are autocorrelated with a lag of 1 (and possibly 4, 5, and 8).

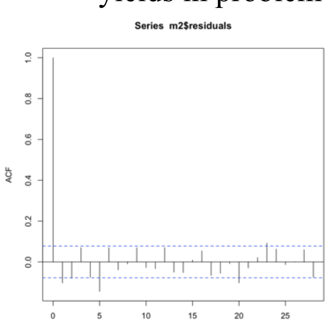
$d3_t = \beta_1 * d1_t + e_t$

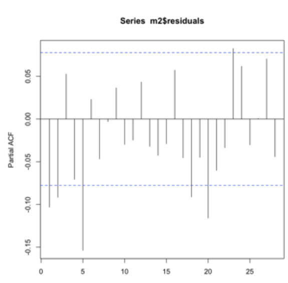
$(y_t - \phi_1 x_t)$ (B) = $\varepsilon_t$ where $y_t = y3_t$ and $x_t = y1_t$

3.3 If we employ 1st differences or as the problem states "monthly change" of yields to predict 3y yield changes with 1y yield changes, then our model to be fitted is as written **above in both equations**. The equations are the same. The first simply regresses one time series $d3_t$ against another $d1_t$. The second employs the backshift notation, but since both are backshift, the B's factor out and 1's cancel out from $(1-B)*y3_t - \phi_1(1-B)*y1_t$

```
Call:
lm(formula = d3t ~ -1 + d1t)

Residuals:
     Min       1Q   Median       3Q      Max 
-0.65669 -0.11132 -0.01054  0.09621  0.81624 

Coefficients:
    Estimate Std. Error t value Pr(>|t|)    
d1t  0.73598    0.01478   49.78   <2e-16 ***
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

Residual standard error: 0.1803 on 634 degrees of freedom
Multiple R-squared:  0.7963,    Adjusted R-squared:  0.796 
F-statistic:  2478 on 1 and 634 DF,  p-value: < 2.2e-16
```

As shown **above** this model produces a significant model and parameter at the 99.9% confidence level but the correlation is less 0.796 than the model employing nominal yields in problem 3.2.






As you can see **above** there is some partial autocorrelation amongst the residuals at lag 5 (and possibly lag 1 and 2).

```
Series: d3t
Regression with ARIMA(0,0,0) errors

Coefficients:
        xreg
      0.7360
s.e.  0.0148

sigma^2 estimated as 0.0325:  log likelihood=187.38
AIC=-370.76   AICc=-370.74   BIC=-361.85
```

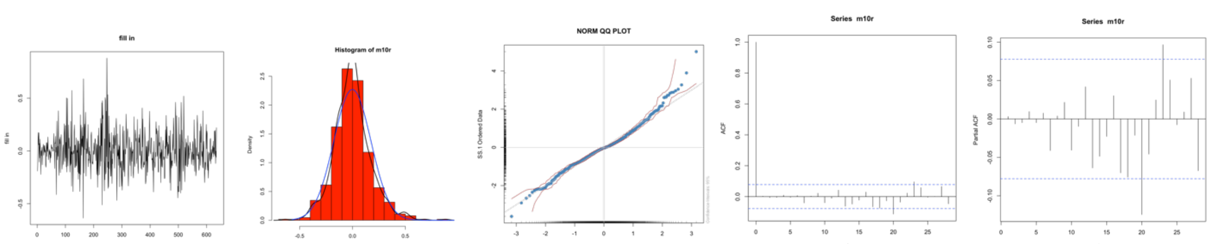
```
Series: d3t
Regression with ARIMA(5,0,0) errors

Coefficients:
          ar1      ar2     ar3      ar4      ar5    xreg
      -0.1178  -0.0845  0.0326  -0.0893  -0.1533  0.7436
s.e.   0.0395   0.0394  0.0396   0.0397   0.0393  0.0142

sigma^2 estimated as 0.03111:  log likelihood=203.7
AIC=-393.4   AICc=-393.23   BIC=-362.23
```

If we regress 1st differences of 3-year yields onto 1st differences of 1y yields we obtain the model on the **LHS above** (matching our parameter estimates in the previous paragraphs of this problem**).** Include autoregression from lag 1 to 5 we obtain the above model on the **RHS above**. From L to R the model AIC improves from -361.85 to -362.23. We can notice that the regression coefficient, xreg, grows in magnitude from 0.7360 to 0.7436 and significance as its t-ratio grows from 49.73 to 52.37. AICc as a second order information criterion would penalize the RHS model for extra parameters but the AICc for LHS is -370.74 vs an improvement to -393.23 on RHS.

3.4 The relationship between 3y and 1y yields' 1st differences can be expressed as follows: For every 1 basis point change in 1y yields, we expect a 0.736 basis point change in 3y yields which is attenuated by the previous 5 days movement in 3y yields, most significantly ar5, ar1, ar4, ar2, and ar3, though ar3 is probably not significant at the 95% confidence level.
Adequacy:

| | n | mean | sd | median | trimmed | min | max | range | skew | kurtosis |
|---|---|---|---|---|---|---|---|---|---|---|
| X1 | 635 | -0.001448499 | 0.1756758 | -0.006843994 | -0.006712762 | -0.6361337 | 0.8784367 | 1.51457 | 0.4529388 | 1.75991 |

```
        One Sample t-test

data:  m10r
t = -0.20778, df = 634, p-value = 0.8355
alternative hypothesis: true mean is not equal to 0
95 percent confidence interval:
 -0.01513848  0.01224149
sample estimates:
   mean of x
-0.001448499
```




The **above** univariate data represent the residuals from the RHS model above with zero mean, moderately positive skew and positive excess kurtosis tested for significance. The time series appears to be stationary with constant mean and variance across the time horizon shown. Aside from the skew and kurtosis, the histogram appears normal. The QQ represents largely normal distribution with 3 outliers on the right extending to nearly 5 SDs. ACF and PACF exhibit no autocorrelation and thus represent independent observations.

```
Title:
 Augmented Dickey-Fuller Test

Test Results:
  PARAMETER:
    Lag Order: 5
  STATISTIC:
    Dickey-Fuller: -10.1143
  P VALUE:
    0.01
```

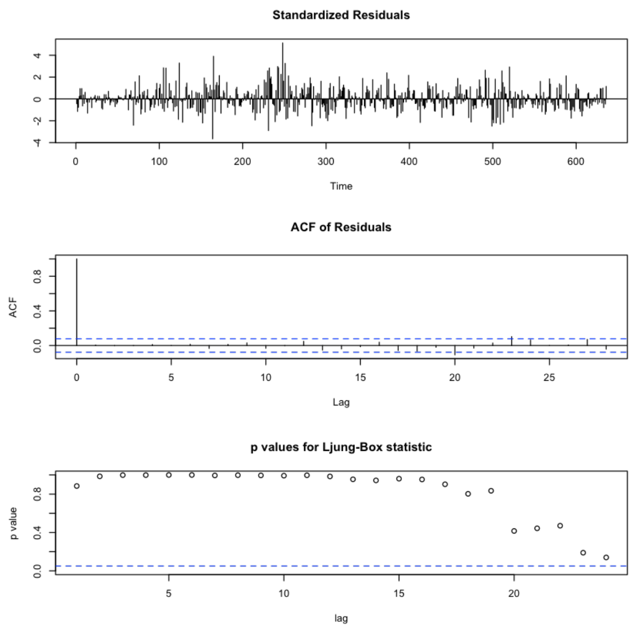
```
#######################
# KPSS Unit Root Test #
#######################

Test is of type: mu with 6 lags.

Value of test-statistic is: 0.1307

Critical value for a significance level of:
                10pct  5pct 2.5pct  1pct
critical values 0.347 0.463  0.574 0.739
```

Both ADF and KPSS **above** represent a time series with no unit root.

$(1-\phi_1 B - \phi_2 B^2 - \phi_3 B^3 - \phi_4 B^4 - \phi_5 B^5 - \phi_6 B^6) y_t = \beta_0 + \beta_1 * x_t + \varepsilon_t$
where $y_t = y3_t$ and $x_t = y1_t$

```
Call:
arima(x = y3t, order = c(6, 0, 0), xreg = y1t)

Coefficients:
         ar1     ar2     ar3      ar4      ar5     ar6  intercept     y1t
      0.8744  0.0329  0.1175  -0.1243  -0.0635  0.1457     1.6424  0.7469
s.e.  0.0395  0.0525  0.0522   0.0527   0.0529  0.0394     0.3652  0.0144

sigma^2 estimated as 0.03063:  log likelihood = 204.46,  aic = -390.92
```

4.1 The **above** equation presents the AR(6) model from our starter code to be fitted and output of that fitting are represented above.




The plots **above** represent a time series with constant mean and variance, an ACF plot that appears to stationary and a plot of Ljung-Box pvalues which suggest independent observations with no autocorrelation.

```
Call:
arima(x = y3t, order = c(6, 0, 0), xreg = y1t, fixed = c2)

Coefficients:
         ar1  ar2     ar3      ar4  ar5     ar6  intercept     y1t
      0.8904    0  0.1332  -0.1553    0  0.1142     1.6336  0.7486
s.e.  0.0295    0  0.0458   0.0458    0  0.0295     0.3622  0.0143

sigma^2 estimated as 0.03071:  log likelihood = 203.63,  aic = -393.25
```

$(1-0.8904*B - 0*B^2 - 0.1332*B^3 - (-0.1553)*B^4 - 0*B^5 - 0.1142*B^6)*y_t = 1.6336 + 0.7486*x_t$

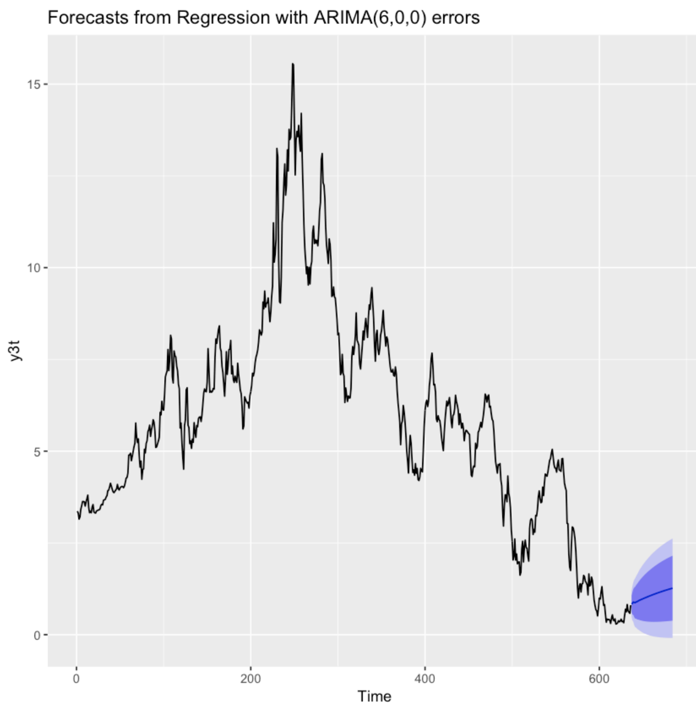
4.2 Zeroing the 2nd & 5th parameter produces the **above** equation for the **above** fitted model. The AIC of the reduced model is more negative -393.25 than the original model -390.92. I would select the reduced model as the timeseries, ACF and Ljung Box plots look similar between the 2 models and we are employing significant predictors in the second model.

4.3 There are 2 real solutions: 1.01281477228878-0i and -1.92368813835684e+00-6e-15i and 4 imaginary solutions:
-0.64251457511639+1.3032647495387i,
-0.64251457511638-1.3032647495387i
1.09795125815042-0.96091871028682i
1.09795125815041+0.96091871028684i

```
Title:
 Augmented Dickey-Fuller Test

Test Results:
  PARAMETER:
    Lag Order: 5
  STATISTIC:
    Dickey-Fuller: -10.1017
  P VALUE:
    0.01
```

```
######################
# KPSS Unit Root Test #
######################

Test is of type: mu with 6 lags.

Value of test-statistic is: 0.3011

Critical value for a significance level of:
                10pct  5pct 2.5pct  1pct
critical values 0.347 0.463  0.574 0.739
```

4.4 The maximum value of the AR(6) model's inverted modulus of roots is 0.98734736830524 which is close to 1.0. The ADF and KPSS tests **above** indicate no presence of unit roots at the 99.9% confidence level for ADF.




5.0 The **above** plot represents the 48-month forecast for 3y yields based upon the model represented in problem 4.2. Forecasts are based upon the mean of the last 48 months 1y yields as employed by the model in problem 4.2. The inner darker purple confidence region is with 95% confidence and the lighter shaded region is with 80% confidence. These do not take into account the uncertainty of the parameters of the model.

fcast <- forecast(m7, xreg=rep(mean(tail(y1t,48)),48))

Figures are shown on the next page for reference:

| | Point Forecast | Lo 80 | Hi 80 | Lo 95 | Hi 95 |
|---|---|---|---|---|---|
| 637 | 0.8362833 | 0.6102840 | 1.062283 | 0.490647102 | 1.181920 |
| 638 | 0.8527046 | 0.5501055 | 1.155304 | 0.389919269 | 1.315490 |
| 639 | 0.8887218 | 0.5370612 | 1.240382 | 0.350903268 | 1.426540 |
| 640 | 0.8809038 | 0.4813816 | 1.280426 | 0.269887264 | 1.491920 |
| 641 | 0.8707632 | 0.4401965 | 1.301330 | 0.212268219 | 1.529258 |
| 642 | 0.8889891 | 0.4375913 | 1.340387 | 0.198635726 | 1.579343 |
| 643 | 0.8978945 | 0.4241080 | 1.371681 | 0.173300640 | 1.622488 |
| 644 | 0.9075623 | 0.4126492 | 1.402475 | 0.150657964 | 1.664467 |
| 645 | 0.9242843 | 0.4096895 | 1.438879 | 0.137279436 | 1.711289 |
| 646 | 0.9366363 | 0.4019284 | 1.471344 | 0.118871200 | 1.754401 |
| 647 | 0.9463807 | 0.3925228 | 1.500239 | 0.099328176 | 1.793433 |
| 648 | 0.9578637 | 0.3864412 | 1.529286 | 0.083948331 | 1.831779 |
| 649 | 0.9681532 | 0.3800512 | 1.556255 | 0.068728807 | 1.867578 |
| 650 | 0.9777983 | 0.3739888 | 1.581608 | 0.054351413 | 1.901245 |
| 651 | 0.9883115 | 0.3697628 | 1.606860 | 0.042322842 | 1.934300 |
| 652 | 0.9986699 | 0.3660281 | 1.631312 | 0.031127793 | 1.966212 |
| 653 | 1.0086921 | 0.3625595 | 1.654825 | 0.020517574 | 1.996867 |
| 654 | 1.0188292 | 0.3598216 | 1.677837 | 0.010964117 | 2.026694 |
| 655 | 1.0287769 | 0.3574270 | 1.700127 | 0.002035911 | 2.055518 |
| 656 | 1.0384616 | 0.3552815 | 1.721642 | -0.006372190 | 2.083295 |
| 657 | 1.0480788 | 0.3535745 | 1.742583 | -0.014073871 | 2.110231 |
| 658 | 1.0575752 | 0.3522088 | 1.762942 | -0.021189589 | 2.136340 |
| 659 | 1.0669200 | 0.3511232 | 1.782717 | -0.027796698 | 2.161637 |
| 660 | 1.0761747 | 0.3503587 | 1.801991 | -0.033865113 | 2.186215 |
| 661 | 1.0853221 | 0.3498695 | 1.820775 | -0.039455548 | 2.210100 |
| 662 | 1.0943425 | 0.3496136 | 1.839071 | -0.044622109 | 2.233307 |
| 663 | 1.1032535 | 0.3495916 | 1.856915 | -0.049372883 | 2.255880 |
| 664 | 1.1120531 | 0.3497833 | 1.874323 | -0.053737941 | 2.277844 |
| 665 | 1.1207359 | 0.3501668 | 1.891305 | -0.057747900 | 2.299220 |
| 666 | 1.1293097 | 0.3507354 | 1.907884 | -0.061416981 | 2.320036 |
| 667 | 1.1377762 | 0.3514767 | 1.924076 | -0.064765187 | 2.340318 |
| 668 | 1.1461345 | 0.3523763 | 1.939893 | -0.067813952 | 2.360083 |
| 669 | 1.1543876 | 0.3534251 | 1.955350 | -0.070578762 | 2.379354 |
| 670 | 1.1625368 | 0.3546132 | 1.970460 | -0.073075719 | 2.398149 |
| 671 | 1.1705826 | 0.3559300 | 1.985235 | -0.075321015 | 2.416486 |
| 672 | 1.1785265 | 0.3573673 | 1.999686 | -0.077328193 | 2.434381 |
| 673 | 1.1863701 | 0.3589170 | 2.013823 | -0.079110198 | 2.451850 |
| 674 | 1.1941142 | 0.3605713 | 2.027657 | -0.080679628 | 2.468908 |
| 675 | 1.2017604 | 0.3623233 | 2.041197 | -0.082047829 | 2.485569 |
| 676 | 1.2093098 | 0.3641664 | 2.054453 | -0.083225508 | 2.501845 |
| 677 | 1.2167638 | 0.3660943 | 2.067433 | -0.084222937 | 2.517750 |
| 678 | 1.2241234 | 0.3681011 | 2.080146 | -0.085049652 | 2.533296 |
| 679 | 1.2313899 | 0.3701815 | 2.092598 | -0.085714634 | 2.548494 |
| 680 | 1.2385645 | 0.3723303 | 2.104799 | -0.086226419 | 2.563355 |
| 681 | 1.2456483 | 0.3745425 | 2.116754 | -0.086593028 | 2.577890 |
| 682 | 1.2526424 | 0.3768137 | 2.128471 | -0.086822027 | 2.592107 |
| 683 | 1.2595481 | 0.3791395 | 2.139957 | -0.086920597 | 2.606017 |
| 684 | 1.2663664 | 0.3815160 | 2.151217 | -0.086895526 | 2.619628 |